DOCUMENT RESUME

ED 270 968 FL 015 701

AUTHOR Ringbom, Hakan
TITLE Word Frequencies in Finnish and Finland-Swedish Learner Language.
PUB DATE 85
NOTE 16p.; In: Ringbom, Hakan, Ed. Foreign Language Learning and Bilingualism; see FL 015 699.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Determiners (Languages); *English (Second Language); *Essays; *Finnish; Foreign Countries; Function Words; Language Usage; Prepositions; Second Language Learning; *Swedish; Vocabulary Development; *Word Frequency; *Writing (Composition)
IDENTIFIERS Finland

ABSTRACT

A comparison of the usage of English articles and prepositions by native speakers of Finnish and of Finland-Swedish studied the word frequencies of these grammatical categories in three levels of English essays written by school-leavers in Finnish-language and Swedish-language schools in Finland. It was hypothesized that the poorest essays written by Finnish-speakers would show lower frequencies in the use of English articles and of prepositions where simple one-to-one correspondences are difficult to establish. The frequencies of these words were expected to rise somewhat in intermediate essays and to rise to almost native-speaker level in the good essays. Because Finland-Swedish is grammatically closer to English, the figures for Finland-Swedish learners were expected to be higher than those of the Finns at the lower levels, but the figures for the best essays were not expected to differ very much. One hundred fifty essays on three levels of proficiency written by Finnish-speaking and the same number written by Swedish-speaking students were chosen to test the hypothesis. The data supported that hypothesis. It is concluded that the pattern of learning articles and prepositions in English where there are no easily established Finnish equivalents represents an avoidance by the students of linguistic features that are absent in the first language and perhaps considered redundant by the learner at early stages. The process should be considered a lack of transfer rather than indirect negative transfer of knowledge from the first language to the second. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Nil Erik Enkvist

Håkan Ringbom

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED270968

WORD FREQUENCIES IN FINNISH AND FINLAND-SWEDISH LEARNER LANGUAGE

Vocabulary learning starts with the learning of high-frequency words, but in order to understand these high-frequency words it is essential to have a knowledge of the functions of the basic grammatical categories to which these words belong. If the learner can rely on at least rough cross-linguistic correspondences between the basic grammatical categories, it will be relatively easy for him to establish simplified equivalences between individual lexical items. When such equivalences can be established for high-frequency words this will help the learner considerably in his learning, whereas, for instance, low-frequency loanwords, formally similar to equivalent L1-words obviously are of much less help. A workable reference frame in the L1 is especially important for the early stages of learning (cf. Ringbom 1985).

Grammatical categories not found in the learner's L1 pose especially great learning problems. In English, the use of the articles has been mentioned by several error analysts as a particularly difficult area to learn for learners whose L1 does not have articles. Thus Oller & Redding found that there was a clear difference between two groups of learners: "G1 (students whose native languages have formal equivalents) performed better on the test of article usage than G2 (students whose native language did not have equivalents....) The differences... were statistically significant" (1971:90 f.). The learning problems for Czech learners was mentioned by Duškova (1969), and for Finnish learners there are several investigations arriving at the same result (Granfors & Palmberg (1976), Herranen (1978), Ringbom (1978a), Sajavaara (1983), and Ekman-Laine (1984)).

215 701

ERIC Full Text Provided by ERIC

BEST COPY AVAILABLE

In particular, Finnish learners, especially at the early and intermediate stages of learning, tend to omit the articles where they should have been included. This obviously reflects a Finn's problems of finding a reference frame for this category. Since function words like articles have only grammatical meaning, not lexical meaning, it is easy for a Finnish learner to perceive articles simply as redundant (cf. George 1972:13 f.), paying little or no attention to them in reception, and neglecting them in production.

When Finnish school-leavers write their matriculation examination they have normally read English for ten years. Having been a national examiner in English for about the same length of time, I have constantly been struck by the enormous differences in quality between the candidates taking part in this compulsory examination, part of which consists of an essay. Although the best of these essays are very native-like, some of the bottom candidates produce English that is hardly comprehensible, even to one who knows the writer's L1. The impression one gets of these is of a kind of reduced code, where especially articles and prepositions are frequently omitted.

A frequency count of the most commonly occurring words in Finnish and Finland-Swedish learner language to verify such subjective impressions may provide an interesting supplement to error analysis, especially when these frequencies are placed beside comparable native speaker figures. The main hypothesis to be tested is, then, that the poorest Finnish essays will show lower frequencies in the use of English articles and of such prepositions where simple one-to-one correspondences are especially difficult to establish. The frequencies of these words will rise in intermediate essays and rise even further, to about native speaker level, in the good essays. The figures for Finland-Swedish learners are expected to be higher than those of the Finns at the low level, but the figures for the best essays are not expected to differ very much either from the corresponding Finnish figures or from the native speaker norm.

In order to test this hypothesis, word frequencies for the most commonly used words were counted. On the basis of the marks awarded by the examiner, 300 essays, 150 written by candidates from Finnish-language schools on ten different topics and the same number from Swedish-language schools on 15 different topics, were selected, representing three

ERIC
BEST COPY AVAILABLE

different levels of proficiency. The 50 good essays from each language group had been awarded a mark of 85 or more out of 99 possible, the 50 intermediate ones were those with marks between 60 and 75 points, and the 50 poor essays had received 50 or less points, i.e. they had either failed (the limit for an accepted essay being 48) or were very close to failure.

The average number of words used in these essays was 201, the total number of words being 29.076 for the Finns and 30.288 for the Swedes.

In fact, it proved difficult to find a sufficient number of the poor category among Swedish speakers. Since much fewer candidates from Swedish schools fail the examination,[1] the upper limit of the worst Swedish essays was raised to 58.

To have only one native-speaker norm would probably have been insufficient and therefore three native-speaker norms were selected as well.

The lowest level consisted of 50 essays written by 15-year-old pupils at Hookergate Comprehensive School, Rowlands Gill, England. These essays were all on the same topic, My School, and their average number of words was 390, the total number of words used 19.540.

The second group was made up of 50 American undergraduates studying in their first year at Purdue University, Indiana. Their essays were on 27 different topics and their average number of words was 706, making up a total of 35.309 words.

The third native-speaker norm is that of the Brown Corpus, Section A, Press Reportage, and the figures are taken from Arne Zettersten's *Word-frequency List* (1978). Each of the 44 samples of this professional writing contained about 2.000 words and the total number of words was 88.753.

The difference in length and the varying number of topics might at first hand seem obstacles to really worthwhile comparisons. However, it should be made clear that this study is only concerned with function words, not content words. Its focus is not content, but language, and the most frequently recurring aspects of language at that.

The fact that number five on the frequency list of the English school children's essays is the word *school* is uninteresting in this context, since it merely reflects the fact that all essays in this group deal with school. What is interesting is to see possible differences in frequencies of function words both within and across the three groups, since this reflects how these different groups use language, not what they write about.

ERIC Full Text Provided by ERIC
EST COPY AVAILABLE

The native-speakers that are, in all respects, most comparable to matriculation examination candidates in Finland as far as general maturity is concerned are the American students, and here the differences in the size and type of samples is negligible if the three proficiency levels are merged into one Finnish and one Swedish category. There may, however, be certain interesting variations in the native-speaker norms (especially school children — university students — professional writers) that may possibly be mirrored in the different Finnish and Finland-Swedish proficiency levels.

Let us first compare the mean frequencies per 100 words of the most commonly used words[2] for the American students with the total number of Finnish and Finland-Swedish (hereafter Swedish) students.

TABLE I

	American students	All Finns	All Swedes
the	5.4	4.4	4.8
to	3.3	3.0	2.7
and	2.4	2.7	2.6
a	2.5	1.9	3.4
of	2.6	2.0	2.1
in	1.7	1.9	2.0
is	1.7	1.6	1.8
it	0.9	1.5	1.5
be	0.7	0.8	1.6
that	1.2	1.6	1.3
have	0.5	1.3	1.4
are	0.6	0.9	0.8
for	0.8	0.7	0.9
but	0.3	0.9	0.7

Table I shows that there seem to be some general differences between learner language and native language. American students make more frequent use of the definite article than Swedes or (especially) Finns, and also the frequency of the word *of* is higher — no doubt mainly because of more frequent use of the *of*-genitive. On the other hand, both Finnish and Swedish learner language uses the words *it*, *have* and *but* more frequently. The high frequency of *but* may be assumed to reflect a general overuse of

ERIC
BEST COPY AVAILABLE

the most common clause connecters, with corresponding infrequent use or no use at all of other clause connecters, whereas native speakers would probably use a wider variety of clause connecters generally. Without detailed studies no immediate explanation can be suggested for the high frequencies of *it* and *have* in both Finnish and Swedish learner language, or for the Swedish high frequency of *be*.[3]

These differences, however, tend to disappear, or are at any rate considerably diminished if we compare the American students only with the best of the Finnish and Swedish candidates. This is true particularly of the frequencies for the definite and indefinite articles and for *of*, as can be seen from Table II.

TABLE II

	American students	Good Finns	Good Swedes
the	5.4	6.0	5.3
to	3.3	3.7	3.0
and	2.4	2.7	2.3
a	2.5	2.7	3.1
of	2.6	2.6	2.8
in	1.7	1.7	1.7
is	1.7	1.7	1.8
it	0.9	1.2	1.3
be	0.7	1.2	2.0
that	1.2	1.1	1.2
have	0.5	1.0	1.2
are	0.6	0.8	0.7
for	0.8	0.8	0.8
but	0.3	0.7	0.8

The examiner's impression of near-native language in these good Finnish and Swedish essays is thus reflected in the similar number of occurrences of high-frequency words. This table shows much smaller differences than table III, where, above all, frequencies of the good group differ very much from those of the poor group, with the intermediate group usually occupying an intermediate position.

ERIC BEST COPY AVAILABLE

TABLE III

Finns

Poor		Intermed.		Good	
the	3.2	the	3.9	the	6.0
and	2.7	and	2.8	to	3.7
to	2.5	to	2.7	and	2.7
in	2.0	in	2.0	a	2.7
it	1.8	that	1.9	of	2.6
is	1.7	of	1.7	in	1.7
that	1.7	a	1.6	is	1.7
have	1.6	is	1.5	it	1.2
of	1.6	it	1.5	be	1.2
a	1.3	have	1.4	that	1.1
was	1.0	was	1.2	have	1.0
are	1.0	but	1.1	are	0.8
but	0.8	are	0.8	for	0.8
at	0.7	for	0.8	but	0.7
not	0.7	at	0.7	has	0.6

Swedes

Poor		Intermed.		Good	
the	4.3	the	4.6	the	5.3
a	3.3	a	4.0	a	3.1
and	2.6	and	2.9	to	3.0
to	2.1	to	2.9	of	2.8
in	1.9	in	2.4	and	2.3
is	1.8	of	1.8	be	2.0
have	1.8	it	1.8	is	1.8
of	1.7	is	1.7	in	1.7
be	1.5	be	1.3	it	1.3
it	1.5	have	1.3	that	1.2
that	1.5	that	1.2	have	1.2
are	1.0	would	1.1	as	0.9
when	0.9	for	1.0	but	0.8
would	0.7	with	0.7	for	0.8
but	0.6	but	0.7	are	0.7

BEST COPY AVAILABLE

School children		Native speakers Students		Brown Corpus	
the	4.9	the	5.4	the	7.4
to	3.3	to	3.3	of	3.3
and	3.1	of	2.6	and	2.5
in	1.9	a	2.5	a	2.5
of	1.9	and	2.4	to	2.5
was	1.8	is	1.7	in	2.3
a	1.6	in	1.7	for	1.1
it	1.1	that	1.2	that	1.0
when	1.0	it	0.9	is	0.9
at	0.9	for	0.8	was	0.8
for	0.7	was	0.7	on	0.8
had	0.7	be	0.7	at	0.7
have	0.7	as	0.7	with	0.7
but	0.7	this	0.7	be	0.6
that	0.6	are	0.6	as	0.6

Table III verifies the hypothesis of the Finns' use of articles. In their article frequencies the best Finns do not differ much either from the Swedes or the Americans, but there is a steadily rising curve from the poor to the good students, with the end points being quite far from one another, as can be seen from Table IV and Diagrams 1–4.

TABLE IV

	Finnish			Swedish			NS		
	Poor	Interm.	Good	Poor	Interm.	Good	School	Univ.	BC
The	3.2	3.9	6.0	4.3	4.6	5.3	4.9	5.4	7.4
A	1.3	1.6	2.7	3.3	4.0	3.1	1.6	2.5	2.5
An	0.11	0.14	0.34	0.35	0.33	0.28	0.09	0.42	0.36
Total	4.6	5.7	9.0	7.9	8.9	8.7	6.6	8.3	10.2

ERIC Full Text Provided by ERIC
BEST COPY AVAILABLE

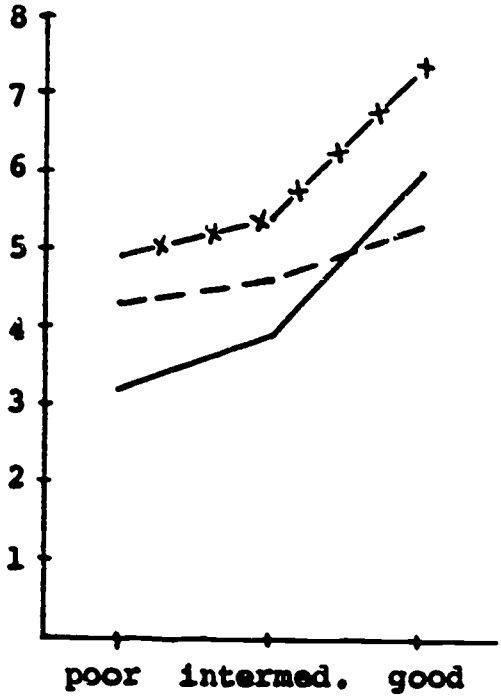

Diagram 1
frequency of *the*

——— = Finns
– – – – = Swedes
—x—x— = Native Speakers

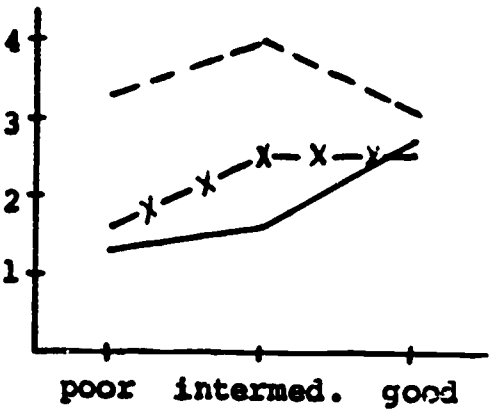

Diagram 2
frequency of *a*

ERIC
Full Text Provided by ERIC
ST COPY AVAILABLE

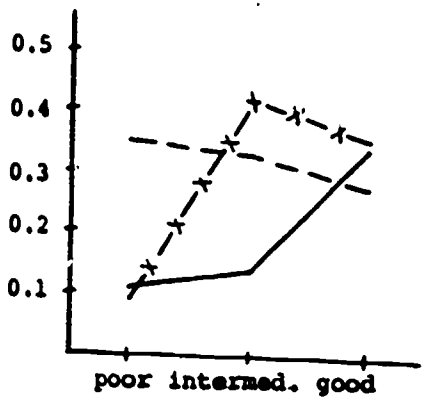

Diagram 3
frequency of *an*

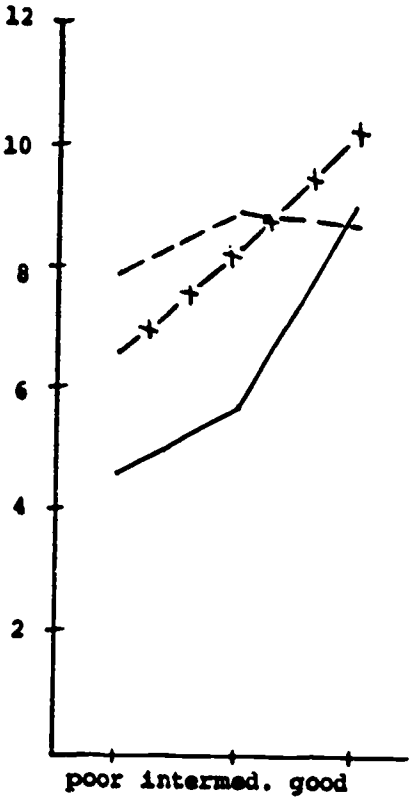

Diagram 4
total frequency of
the, a and *an*

ERIC
BEST COPY AVAILABLE

In addition to articles, English prepositions make up another main problem area for Finnish learners, since this is a word class that hardly exists in Finnish, and correspondences to Finnish cases are much less naturally established than to, for instance, the prepositions in Swedish. However, not all prepositions pose the same problems for the Finnish learner. For certain prepositions, notably *in* and *with*, Finnish equivalents (the inessive case for *in* and the postposition *kanssa* for *with*) can be established that will work in most cases. Those English prepositions, on the other hand, which are so multifunctional that a Finnish correspondence is not easily established are, above all, *of, on* and by, for which a very wide variety of Finnish constructions will be used. With these prepositions one might therefore expect the same picture to emerge as for the articles: that the Finns at the low and intermediate proficiency level use these words much less frequently than either the American students or the corresponding Swedish learners, whereas the figures would be by and large similar for *in* and *with*.

TABLE V

	Finnish			Swedish			NS		
	Poor	Interm.	Good	Poor	Int.	Good	School	Univ.	BC
In	2.0	2.0	1.7	1.9	2.4	1.7	1.9	1.7	2.3
With	0.4	0.5	0.5	0.5	0.7	0.5	0.5	0.6	0.7
Of	1.5	1.7	2.6	1.7	1.8	2.8	1.9	2.6	3.3
On	0.2	0.3	0.4	0.4	0.5	0.5	0.6	0.6	0.7
By	0.16	0.19	0.31	0.06	0.18	0.26	0.18	0.5	0.6

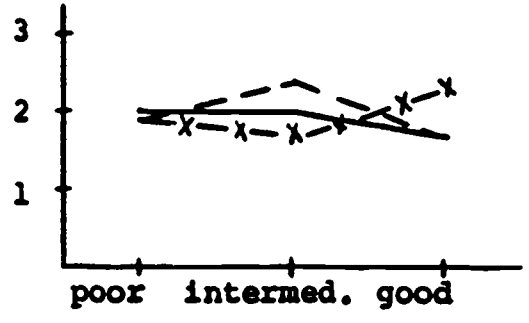

Diagram 5
frequency of *in*

ERIC
BEST COPY AVAILABLE

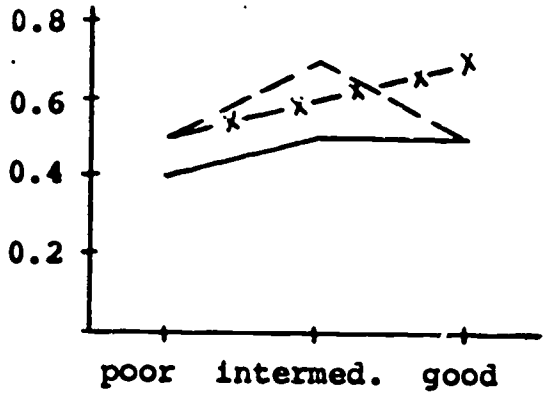

Diagram 6
frequency of *with*

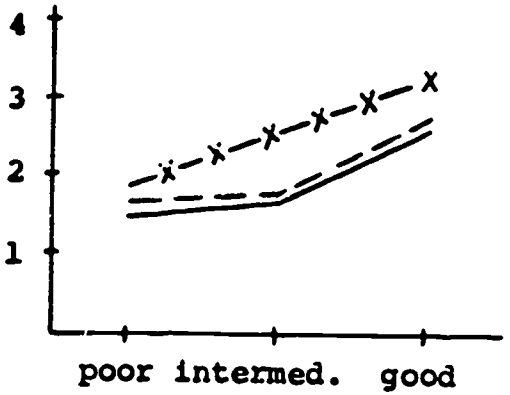

Diagram 7
frequency of *of*

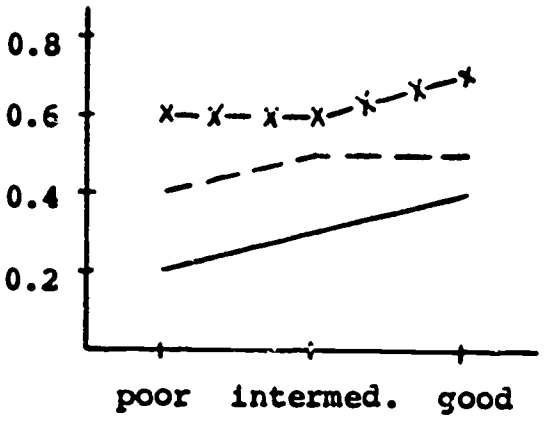

Diagram 8
frequency of *on*

BEST COPY AVAILABLE

Diagram 9
frequency of *by*

On the whole, the figures in Table V conform to the expectation. It is, however, dangerous to overemphasize the contrastive linguistic aspects here, since the table also reveals other differences. In several respects, especially the poor and intermediate levels of both Finnish and Swedish learners have frequencies very similar to those of the English schoolchildren. This is true particularly of the word *of*, where sparse use of the *of*-genitive must obviously be a main reason. The same is true of the word *by*, where the use of the agentive *by* will provide at least a partial explanation. *Of*-genitives and especially passives tend to be used primarily by writers of a certain linguistic sophistication: i.e., native speakers at university level and the really good foreign learners. Further investigation of the context is, however, needed here, since we need to know also whether the words are correctly used or not, but the tendency is the same as that of Danish learners. According to Faerch (1981) *in* was much more frequent than *of* in Danish schoolchildren, but less frequent than *of* in the English of university students reading English. It is not so much the decreasing use of *in* (26.5% vs 21.4% of all prepositions), as the increasing use of *of* (14.8% vs 27.1%) that causes the difference. Similarly, the use of *by* increases from 0.2% of all prepositions used by Danish schoolchildren to 5.4% of the prepositions used by university students. It is thus easy to agree with Faerch and discern the same pattern in both Finnish and Finland-Swedish learner language when he states (1981:15): "Even without a supplementary qualitative analysis of the functions of these prepositions I think it is safe to interpret these results as primarily indicating the growth of *syntactic* structures (*of*-genitives and passives)."

ERIC
BEST COPY AVAILABLE

* * *

The differences between the frequencies of Finnish and Finland-Swedish learners are particularly illuminating about the Finnish learners. Above all, they pinpoint the differences between successful learners and unsuccessful learners who have had the same amount of training. There is a clear progression in the use of such English forms without easily establishable equivalents in Finnish: from a very low frequency in the poor Finnish candidates, who obviously tend to omit these words when they should have been used, to near-native frequency in the good candidates. It is important to note this avoidance of the use of linguistic features absent in the L1 and therefore easily perceived as redundant at the earliest stages of learning and perhaps as merely 'difficult' at subsequent stages of learning. It is a clear indication of the important role of the mother tongue in such learners as are very far from native-like proficiency, in spite of having got quite extensive training. The term (negative) transfer or interference would, however, not be a very happy term to use in this context, since it is difficult to see exactly what has been transferred. Instead of classifying this kind of avoidance as a case of indirect negative transfer or interference (Dagut & Laufer 1982), we should take it as an example of lack of (positive) transfer, since it is, in fact, the absence of a reference frame that has caused this avoidance. This tallies well with Schachter's (1974) findings that Chinese learners of English avoid producing relative clauses, since they find them quite difficult, not having such construction in their L1.[4]

NOTES

1 No exact figures are available here to support this statement, but the following table for the last eleven years shows the mean score of candidates from Finnish and Swedish schools in two other parts of the English examination, a listening comprehension test and a reading comprehension test:

ERIC
BEST COPY AVAILABLE

TABLE VI

	LC (max. 30 p)		RC (max. 30 p)	
	Finnish	Swedish	Finnish	Swedish
1974	19.7	22.4	24.1	25.7
1975	21.6	24.8	22.8	24.7
1976	18.5	22.4	18.6	22.0
1977	18.5	23.2	23.4	25.3
1978	22.1	25.2	22.9	25.2
1979	23.3	26.4	19.4	21.6
1980	20.9	25.0	18.4	19.9
1981	22.5	25.4	22.1	25.0
1982	21.2	24.6	23.6	26.4
1983	22.8	26.1	23.3	25.4
1984	25.6	28.3	22.3	24.4

Number of candidates: Approximately 11 x 25.000 Fi. =275.000
11 x 1.500 Sw.= 16.500

For a discussion of the consistently greater difference between Finns and Swedes in the listening comprehension test than in the reading comprehension test, see Ringbom 1978b.

2 The frequencies of personal and possessive pronouns are highly variable, depending primarily not on the writer's use of language, but on his approach to the topic, and the words *he* and *I*, at least one of which normally occurs in any top twenty list, have therefore not been included in my tables of high-frequency words.

3 A more detailed analysis of learner language would, of course, have to distinguish between the different uses of some of these high-frequency words, especially *that* and *to*. The frequency of the latter interestingly shows a steady rise from poor to good proficiency level in both Finnish and Swedish learners. A possible explanation of this will, however, have to be postponed to another context.

4 For help with various tasks in connection with this paper I am grateful to the following: Monica Andersson, Sonja Granbacka, Auli Haapaniemi, Seppo Hovila, Jari Lehtonen, Sari Monni and Gerd Nyholm.

REFERENCES

Dagut, Menachem & Batia Laufer. 1982. How Intralingual Are "Intralingual Errors?" *Error Analysis. Contrastive Linguistics and Second Language Learning* (Special Issue of *IRAL*). Papers from the 6th International Congress of Applied Linguistics, Lund 1981, edited by Gerhard Nickel and Dietrich Nehls, 19-41. Heidelberg: Julius Groos.

ERIC
BEST COPY AVAILABLE

Duškova, L. 1969. On Sources of Errors in Foreign Language Learning. *IRAL* 7, 11–36.

Ekman-Laine, Ulla 1984. An Analysis of Three Main Types of Errors Made by Finns in the Matriculation Examination in the Autumn of 1978. Unpublished M.A. thesis, Åbo Akademi

Faerch, Claus. 1981. Some Problems in Describing IL vocabulary-discussion paper. Unpublished paper written for AILA 1981, Interlanguage Workshop.

George, H.V. 1972. *Common Errors in Language Learning. Insights from English.* Rowley, Mass: Newbury House.

Granfors, Tom & Rolf Palmberg. 1976. Errors Made by Finns and Swedish-speaking Finns Learning English at a Commercial-College Level. *Errors Made by Finns in the Learning of English* edited by Håkan Ringbom & Rolf Palmberg. AFTIL 5, Dept. of English, Åbo Akademi

Herranen, Tauno. 1978. Errors Made by Finnish University Students in the Use of the English Article System. *Jyväskylä Contrastive Studies 6: Further Contrastive Papers* by Kari Sajavaara, Jaakko Lehtonen and Raija Markkanen (eds.), 74–95.

Oller, J.W. Jr & E. Redding. 1971. Article Usage and Other Language Skills. *Language Learning* 21:1, 85–95.

Ringbom, Håkan. 1978a. What Differences are there between Finns and Swedish-speaking Finns Learning English? *Papers and Studies in Contrastive Linguistics* (Poznań), 133–145.

Ringbom, Håkan. 1978b. On the Difficulty of English Listening Comprehension in Finland. *Tempus* 1978:2, 6–9.

Ringbom, Håkan. 1985. Transfer in relation to some other Variables in L2-Learning. *Foreign Language Learning and Bilingualism* by Håkan Ringbom (ed.). Publications of the Research Institute of the Åbo Akademi Foundation vol. 105, 9–21.

Sajavaara, Kari. 1983 The Article Errors of Finnish Learners of English, *Finnish-English Language Contact: Papers from a Workshop* edited by Claes-Christian Elert & Aimo Seppanen. Umeå Papers in English 4, 72–87.

Schachter, Jacqueline (1974). An Error in Error Analysis. *Language Learning* 24, 205–213.

ERIC ST COPY AVAILABLE